Virus Genes 8:3, 243–247, 1994

Predicted Stem-Loop Structures and Variation in Nucleotide Sequence of 3′ Noncoding Regions Among Animal Calicivirus Genomes

BRUCE S. SEAL,[1] JOHN D. NEILL,[2] & JULIA F. RIDPATH[2]

Virology Swine Research Unit[1] and Virology Cattle Research Unit,[2] National Animal Disease Center, USDA, Agriculture Research Service, Ames, IA, USA

Received and accepted November 30, 1993

Requests for reprints should be addressed to Dr. Bruce S. Seal at his present address: Southeast Poultry Research Laboratory, USDA, ARS, P.O. Box 5657, 934 College Station Rd., Athens, GA 30604.

Abstract. Caliciviruses are nonenveloped with a polyadenylated genome of approximately 7.6 kb and a single capsid protein. The "RNA Fold" computer program was used to analyze 3′-terminal noncoding sequences of five feline calicivirus (FCV), rabbit hemorrhagic disease virus (RHDV), and two San Miguel sea lion virus (SMSV) isolates. The FCV 3′-terminal sequences are 40–46 nucleotides in length and 72–91% similar. The FCV sequences were predicted to contain two possible duplex structures and one stem-loop structure with free energies of −2.1 to −18.2 kcal/mole. The RHDV genomic 3′-terminal RNA sequences are 54 nucleotides in length and share 49% sequence similarity to homologous regions of the FCV genome. The RHDV sequence was predicted to form two duplex structures in the 3′-terminal noncoding region with a single stem-loop structure, resembling that of FCV. In contrast, the SMSV 1 and 4 genomic 3′-terminal noncoding sequences were 185 and 182 nucleotides in length, respectively. Ten possible duplex structures were predicted with an average structural free energy of −35 kcal/mole. Sequence similarity between the two SMSV isolates was 75%. Furthermore, extensive cloverleaflike structures are predicted in the 3′ noncoding region of the SMSV genome, in contrast to the predicted single stem-loop structures of FCV or RHDV.

Key words: recombinant DNA, graphics modeling, RNA folding, positive-strand RNA virus, nucleotide sequencing

The caliciviridae are a group of relatively uncharacterized animal viruses that have positive sense RNA genomes approximately 7.6 kb in size (1–3). These nonenveloped viruses have a density of 1.35 g/cc, a single capsid protein ranging from 60 to 70 kD, and a Vpg protein covalently linked at the 5′ end of the genomic RNA (2,4,5). Members of this group include feline caliciviruses (FCV), San Miguel sea lion virus (SMSV), vesicular exanthema of swine virus (VESV), rabbit hemorrhagic disease virus (RHDV), and the Norwalk virus (6). Candidate caliciviruses have also been isolated from chickens (7), dogs (8), and cattle (9). Caliciviruses characteristically have three open reading frames in their genomes. The nonstructural proteins are encoded in the 5′ half of the genome and include a 2C polypeptide, a cysteine protease, and an RNA-dependent RNA polymerase (1,10). The structural capsid gene is located at the beginning of the 3′ half of the genome followed by a small putative DNA binding protein gene with an untranslated region directly preceeding the Poly-A tail (1,11,12).

Disclaimer: No endorsements are herein implied. Brand names are necessary to report factually on available data; however, the USDA neither guarantees nor warrants the standards of the products, and the use of the names by the USDA implies no approval of the products to the exclusion of others that may also be suitable.

The 3′-terminal sequences of several positive-strand RNA viruses have been reported and analyzed for secondary structures in the untranslated region. Possible secondary structures were first predicted in the 3′-terminal regions of turnip yellow mosaic virus (TYMV) genomic RNA (13,14), and structural models have been proposed for TYMV (15–17). Among animal viruses the 3′ terminal untranslated region of several alphavirus genomes have been compared and demonstrated to contain blocks of repeated sequences (18–21). Stable secondary structures have also been detected in the 3′-terminal untranslated regions of numerous flavivirus genomes (22–27). Most recently, secondary structures have been detected in poliovirus 3′-terminal untranslated region and are thought to be important in viral RNA amplification during infection (28). The purpose of the present communication is to report and compare the 3′-untranslated regions of several animal calicivirus genomes and to analyze these regions for predicted secondary structures that may be present.

Sequences of the 3′-terminal untranslated genomic region of five FCV isolates have been compared in Fig. 1A. These sequence areas are 40–45 nucleotides in length among the five isolates and vary in sequence similarity by 78–91%. The NADC isolate was obtained most recently in 1989 (38) and was 45 nucleotides in length. The KCD isolate was made in 1957 (34) and contained the smallest number of nucleotides in this region of the genome (Fig. 1A). Conserved sequences exist among the five isolates, consisting of a CCCTT and CGC in the interior area of this region of feline calicivirus genomes. The 12 nucleotides just prior to the polyA tail composed of the sequence CTAACCCCAGGG are also conserved among all five isolates. Immediately upstream from this sequence is a GCG island that is present in all but the KCD isolate (Fig. 1A). The KCD isolate has been maintained

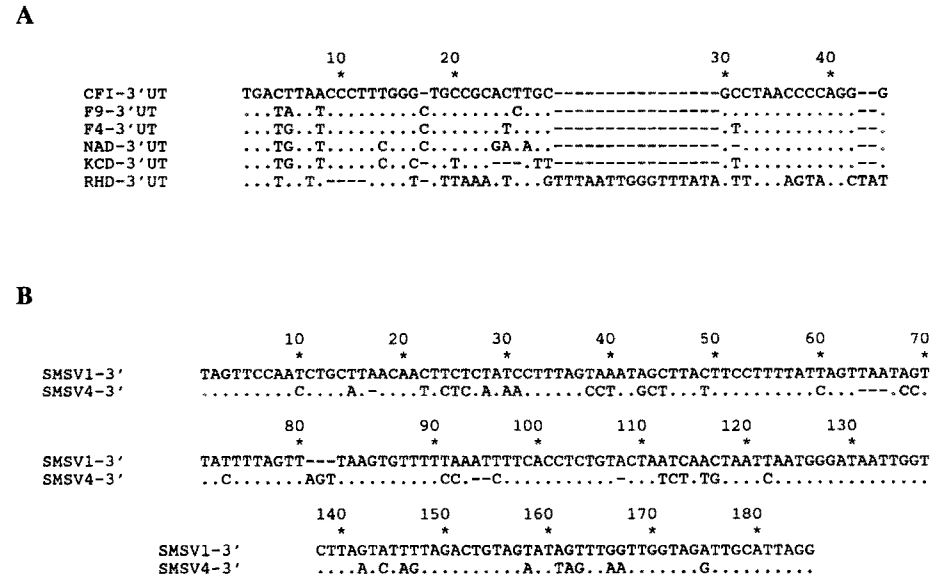

Fig. 1. Nucleotide sequence comparison of the 3′-noncoding region among animal calicivirus RNA genomes. *A:* Nucleotide sequence alignment of the genomic RNA 3′-noncoding region among isolates of feline calicivirus and rabbit hemorrhagic disease virus. *B:* Nucleotide sequence alignment of the genomic RNA 3′-noncoding region between two isolates of San Miguel sea lion virus. Published sequences for the CFI/68 (11), F4 (33), and F9 (1) viruses were aligned with two more isolates of feline calicivirus, NADC and KCD, and rabbit hemorrhagic disease virus (RHDV; 30). The NADC virus isolation was made from a healthy cat at the Ames, Iowa animal shelter, and the KCD isolate was a standard laboratory strain (34) obtained from the American Type Culture Collection. Nucleotide sequencing (35,36) of the NADC and KCD isolates was from cDNA obtained by unidirectional cloning (37) of genomic RNA. Genbank accession numbers are L09718 for the NADC isolate and L09719 for the KCD isolate. Sequences for the SMSV 1 and SMSV 4 isolates were from published data (12). Genomic RNA sequences were compared using the ALIGN (Scientific & Educational Software) computer program.

longer in cell culture than any of the other FCV isolates, and this may have contributed to the reduced number of bases in this region of the genome.

To determine what potential secondary structures may occur in the 3′-terminal untranslated region, these sequences were subjected to computer analysis (29). The FCV sequences contained one predicted stem-loop in their 40–45 base untranslated terminal region (Fig. 2). The CFI/68 isolate contains one duplex that pairs bases 12–15 with bases 46–43 and a second duplex pairing bases 16–23 with bases 38–31. A total energy prediction for the CFI/68 isolate's 3′-terminal structure was −17.9 kcal/mol. The KCD isolate also contained two predicted duplex structures, while three such structures were predicted for the NADC isolate (Fig. 2). A total energy of −5.8 kcal/mol was predicted for the KCD isolate and −18.2 kcal/mol was predicted for the NADC isolate in their 3′-terminal untranslated regions. Analysis of the 3′-terminal untranslated region of the FCV and F9 isolated produced similar results (Fig. 2). Both isolates contained two duplex structures in this region of their genome with total energy predictions of −2.1 kcals/mol. The predicted structures of the FCV genomic 3′-terminal untranslated region resemble single stem-loop structures found in genomic RNA of a poliovirus mutant able to replicate only at 39.5°C and not at 37°C (28). All the FCV isolates exhibited equivalent replication

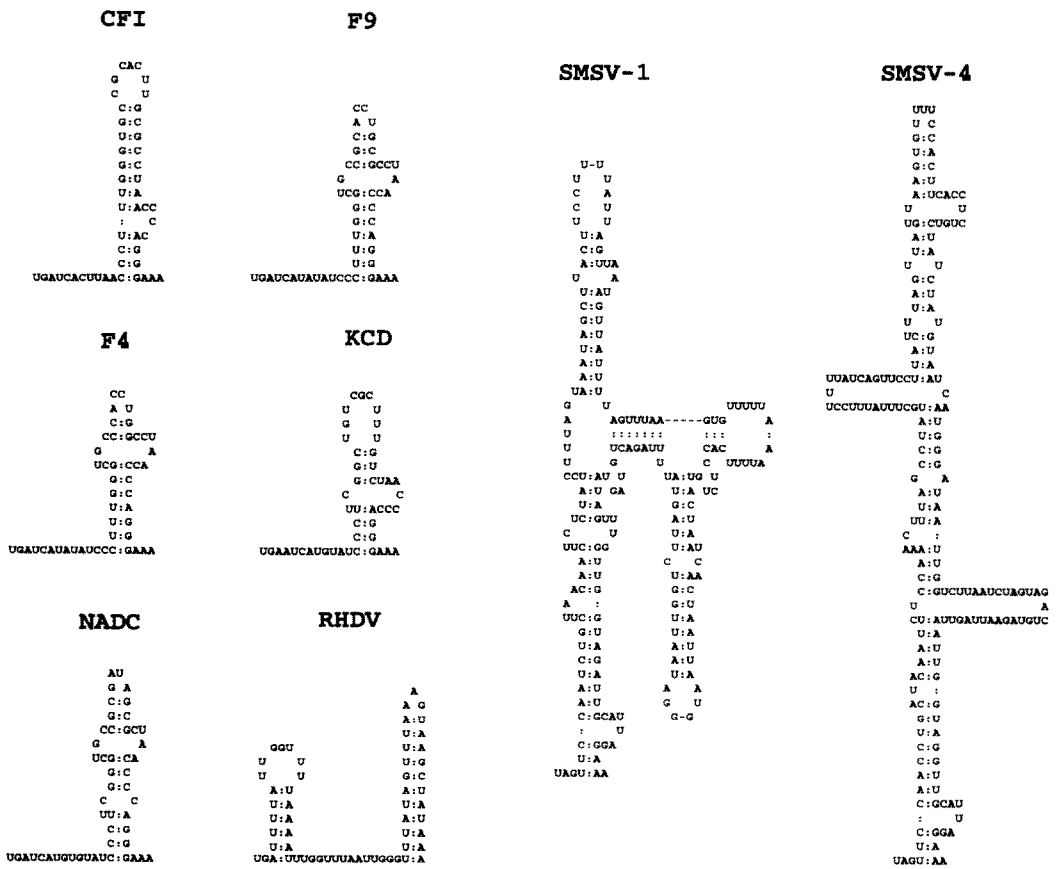

Fig. 2. Predicted secondary structures for the 3′-noncoding regions of animal calicivirus genomic RNA. Feline calicivirus isolates CFI/68, F4, NADC, F9, KCD; rabbit hemorrhagic disease virus (RHDV), and San Miguel sea lion virus isolates 1 and 4 (SMSV-1 and SMSV-4) were analyzed by the RNA Fold computer program (Scientific & Educational Software) for the presence of stem-loop structures in the 3′-noncoding region by the method of Martinez (29). For all RNA folding analyses, a portion of the poly A tail was included, since it has been determined that these adenine residues may base pair with pyrimidines in the 3′ terminus of polyadenylated viral genomes (28).

curves at 37°C. Hence, the secondary structures present in the FCV 3′-terminal untranslated region are sufficient for viral replication.

The other positive-strand RNA animal virus considered to be a member of the caliciviridae family is the rabbit hemorrhagic disease virus (RHDV) that infects the liver of its host. This virus has 3′-terminal noncoding sequences of 54 nucleotides in length (2) that is 100% similar among the isolates sequenced to date (30). These regions of their genomes are approximately 49% similar to the FCV isolates 3′-noncoding sequences, and the stop codon UGA is utilized by RHDV, as is the case for FCV isolates (Fig. 1A). Two stem-loop structures are predicted to occur with a calculated free energy of −7.1 kcal/mol (Fig. 2). The structures predicted to occur in the RHDV genomic 3′-terminal untranslated region also resemble similar structures found in wild-type poliovirus and coxsackievirus B1 genomic RNA that presumably interact with the viral polymerase during viral RNA amplification (28).

Another calicivirus group, the SMSV isolates, were also compared for sequence similarity among their 3′-terminal untranslated genomic region (Fig. 1B). The SMSV isolates have terminal untranslated regions that are 185 and 182 nucleotides long in type 1 and 4 isolates, respectively. This region of the genome is 75% similar between these two isolates and contains islands of sequence similarity separated by areas of one to three base differences. The SMSV isolates have the stop codon UAG (Fig. 1B), different from the UGA utilized by FCV and RHDV (Fig. 1A). Similar to the situation in FCV, there is a region of conserved sequences directly 5′ to the polyA tail. However, the sequence of 10 nucleotides, TTGCATTAGG, in SMSV 1 and 4 has no similarities to the FCV sequences directly preceding the polyA tail.

Potential secondary structures in the 3′-untranslated region of the SMSV isolates are much more extensive than in the FCV isolates because of the increased number of nucleotides in this area of the genome (Fig. 2). A more complex cloverleaf structure was predicted for both SMSV isolates. The SMSV-1 isolate has 10 duplex structures present, with a total energy of −32.6 kcal/mol, while the SMSV-4 isolate also has 10 duplex structures, with a total energy of −38.1 kcal/mol. The secondary structures predicted to occur in the 3′-terminal untranslated region of SMSV genomic RNA resemble structures predicted to occur in homologous regions of various positive-stranded plant virus genomes (14,16,17,23,31) and Dengue virus type 2 RNA (25). Similar secondary structures of flaviviruses (22,24) and encephalitis viruses (26) have been proposed to be important for replication of viral RNA.

No sequence similarities to the calicivirus 3′-terminal genomic region were found in the EMBL/GenBank database. Specifically, no apparent sequence similarities were detected to the alphaviruses, flaviviruses, or to any of the positive-strand RNA plant viruses. Although the 3′-terminal untranslated regions of calicivirus genomes are quite variable in their sequence length, they all may form stem-loop structures. These areas of the genome are believed to be important in providing replication signals responsible for amplification of RNA molecules (32) and may bind proteins involved with tRNA metabolism (16,17). Consequently, the terminal stem-loop structures of caliciviruses may play a role in the viral replication complex, as has been proposed for poliovirus (28). The two better characterized caliciviruses, FCV and SMSV, are very different in this region of the genome. The FCV isolates have a short 43–48 nucleotide untranslated terminal region, while the SMSV isolates have a much longer 181 or 185 nucleotide terminal untranslated region with a more extensive cloverleaf structure. The RHDV viruses have a terminal untranslated region intermediate in length to the FCV and SMSV isolates.

Acknowledgments

The authors thank Ms. Deborah Floan for assistance with the primary nucleotide sequencing and Ms. Judy Lundy of the Iowa State University Nucleic Acids Facility for oligonucleotide synthesis. The authors also thank Mr. Gene Hedberg for graphics illustration and Ms. Sue Ohlendorf for administrative assistance. The authors thank Dr. R. Zeurner, Dr. S. Halling, and Dr. M. Perdue for helpful comments.

References

1. Carter M.J., Milton I.D., Meanger J., Bennet M., Gaskel R.M., and Turner P.C., Virology *190*, 443–448, 1992.
2. Meyers G., Wirblich C., and Theil H.J., Virology *184*, 664–676, 1991.
3. Meyers G., Wirblich C., and Theil H.J., Virology *184*, 677–686, 1991.
4. Bachrach H.L. and Hess W.R., Biophys Res Commun *55*, 141–149, 1973.
5. Burroughs J.N. and Brown F., J Gen Virol *41*, 443–446, 1978.
6. Schaffer F.L., Fraenkel-Conrat H., and Wagner R., (eds). *Comparative Virology, 14*. Plenum Press, New York, 1979, pp. 249–284.
7. Cubitt W.D. and Barret A.D., J Gen Virol *66*, 1431–1438, 1985.
8. Schaffer F.L., Soergal M.E., Black J.W., Skilling D.E., Smith A.W., and Cubitt W.D., Arch Virol *84*, 181–195, 1985.
9. Smith A.W., Mattson D.E., Skilling D.E., and McKeirnan A.J., Am J Vet Res *44*, 851–855, 1983.
10. Neill J.D., Virus Res *17*, 145–160, 1990.
11. Neill J.D., Reardon I.M., and Heinrikson R.L., J Virol *65*, 5440–5447, 1991.
12. Neill J.D., Virus Res *24*, 211–222, 1992.
13. Pinck M., Yot P., Chapeville F., and Duranton H.M., Nature *226*, 954–956, 1970.
14. Yot P., Pinck M., Haenni A.L., Duranton H.M., and Chapeville F., Proc Natl Acad Sci USA *67*, 1345–1352, 1970.
15. Mans R.M.W., Van Steeg M.H., Verlaan P.W.G., Pleij C.W.A., and Bosch L., J Mol Biol *223*, 221–232, 1992.
16. Rietveld K., Linschooten K., Pleij C.W.A., and Bosch L., EMBO J *3*, 2613–2619, 1984.
17. Rietveld K., Pleij C.W.A., and Bosch L., EMBO J *2*, 1079–1085, 1983.
18. Dalgarno L., Rice C.M., and Strauss J.H., Virology *129*, 170–187, 1983.
19. Faragher S.G. and Dalgarno L., J Mol Biol *190*, 141–148, 1986.
20. Ou J-H., Strauss E.G., and Strauss J.H., Virology *109*, 281–289, 1981.
21. Ou J-H., Trent D.W., and Strauss J.H., J Mol Biol *156*, 19–730, 1982.
22. Brinton M.A., Fernandez A.V., and Dispoto J.H., Virology *153*, 113–121, 1986.
23. Grange T., Boulouy M., and Girard M., FEBS Lett *188*, 159–163, 1985.
24. Hahn C.S., Hahn Y.S., Rice C.M., Lee E., Dalgarno L., Strauss E.G., and Strauss J.H., J Mol Biol *198*, 33–41, 1987.
25. Mohan M. and Padmanabhan R., Gene *108*, 185–191, 1991.
26. Takegami T., Washizu M., and Yasui K., Virology *152*, 483–486, 1983.
27. Wengler G. and Castle E., J Gen Virol *67*, 1183–1188, 1986.
28. Jacobson S.J., Konings D.A.M., and Sarnow P., J Virol *67*, 2961–2971, 1993.
29. Martinez H.M., Nucleic Acids Res *12*, 323–334, 1984.
30. Milton I.D., Vlasak R., Nowotny N., Rodak L., and Carter M.J., FEMS Microbiol Lett *93*, 37–42, 1992.
31. Florentz C. and Giegé R., J Mol Biol *191*, 117–130, 1986.
32. Weiner A.M. and Maizels N., Proc Natl Acad Sci USA *84*, 7383–7387, 1987.
33. Tohya Y., Taniguchi Y., Takahashi E., Utagawa E., Takeda N., Miyamura K., Yamazaki S., and Mikami T., Arch Virol *183*, 810–814, 1991.
34. Fastier L.B., Am J Vet Res *18*, 382–389, 1957.
35. Prober J.M., Trainor G.L., Dan R.J., Hobbs F.W., Robertson C.W., Zaqursky R.J., Cocuzza A.J., Jensen M.A., and Burmeister K., Science *238*, 336–341, 1987.
36. Sanger F., Nickles S., and Carlson A.R., Proc Natl Acad Sci USA *74*, 5463–5467, 1977.
37. D'Alessio J.M. and Gerard G.F., Nucleic Acids Res *16*, 1999–2014, 1988.
38. Seal B.S., Ridpath J.F., and Mengeling W.L., J Gen Virol *74*, 2519–2524, 1993.